Rev. CENIC Cienc. Biol.; Vol. 53. (3): 219-242. Año. 2022. e-ISSN: 2221-2450.

ARTICULO DE INVESTIGACION

Finding electrophysiological sources of aging-related processes using penalized least squares with Modified Newton-Raphson algorithm

Encontrar fuentes electrofisiológicas de procesos relacionados con el envejecimiento utilizando mínimos cuadrados penalizados con el algoritmo de Newton-Raphson modificado

Mayrim Vega Hernández [a] (0000-0003-0817-7188).
Darío Palmero Ledón [a] (0000-0001-8890-5987).
José M. Sánchez-Bornot [b] (0000-0003-4014-4255).
Jhoanna Pérez Hidalgo Gato [a] (0000-0001-7647-0779).
Daysi García Agustin [a] (0000-0002-2296-1078).
Pedro A. Valdés Sosa [c] (0000-0001-5485-2661).
Eduardo Martínez Montes [a,*] (0000-0002-5852-964X).

[a] Cuban Center for Neurosciences, Havana, Cuba.
[b] School of Computing and Intelligent Systems, Ulster University, UK.
[c] The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China.

*eduardo@cneuro.edu.cu

Recibido: 22 de junio de 2022; Aceptado: 12 de julio de 2022;

ABSTRACT

In this work, we evaluate the flexibility of a modified Newton-Raphson (MNR) algorithm for finding electrophysiological sources in both simulated and real data, and then apply it to different penalized models in order to compare the sources of the EEG theta rhythm in two groups of elderly subjects with different levels of declined physical performance. As a first goal, we propose the MNR algorithm for estimating general multiple penalized least squares (MPLS) models and show that it is capable to find solutions that are simultaneously sparse and smooth. This algorithm allowed to address known and novel models such as the Smooth Non-negative Garrote and the Non-negative Smooth LASSO. We test its ability to solve the EEG inverse problem with multiple penalties -using simulated data- in terms of localization error, blurring and visibility, as compared with traditional algorithms. As a second goal, we explore the electrophysiological sources of the theta activity extracted from resting-state EEG recorded in two groups of older adults, which belong to a longitudinal study to assess the relationship between measures of physical performance (gait speed) decline and normal cognition. The groups contained subjects with good and bad physical performance in the two evaluations (6 years apart). In accordance to clinical studies, we found differences in EEG theta sources for the two groups, specifically, subjects with declined physical performance presented decreased temporal sources while increased prefrontal sources that seem to reflect compensating mechanisms to ensure a stable walking.

Keywords: Aging-related disorders; electrophysiological sources; modified MNR; inverse problem; multiple penalized least squares; gait speed disorder.

RESUMEN

En este trabajo, evaluamos la flexibilidad de un algoritmo modificado de Newton-Raphson (MNR) para encontrar fuentes electrofisiológicas en datos reales y simulados, y luego lo aplicamos a diferentes modelos penalizados para comparar las fuentes del ritmo theta de EEG en dos grupos de sujetos de edad avanzada con diferentes niveles de disminución del rendimiento físico. Como primer objetivo, proponemos el algoritmo MNR para estimar modelos generales de mínimos cuadrados penalizados múltiples (MPLS) y demostramos que es capaz de encontrar soluciones que son simultáneamente dispersas y suaves. Este algoritmo permitió abordar modelos conocidos y novedosos como el Garrote Liso No negativo y el LASSO Liso No negativo. Probamos su capacidad para resolver el problema inverso del EEG con múltiples penalizaciones -utilizando datos simulados- en términos de error de localización, desenfoque y visibilidad, en comparación con los algoritmos tradicionales. Como segundo objetivo, exploramos las fuentes electrofisiológicas de la actividad theta extraída del EEG en estado de reposo registrado en dos grupos de adultos mayores, que pertenecen a un estudio longitudinal para evaluar la relación entre las medidas de disminución del rendimiento físico (velocidad de la marcha) y la normalidad. cognición. Los grupos contenían sujetos con buen y mal desempeño físico en las dos evaluaciones (6 años de diferencia). De acuerdo con los estudios clínicos, encontramos diferencias en las fuentes theta de EEG para los dos grupos, específicamente, los sujetos con un rendimiento físico disminuido presentaron fuentes temporales disminuidas mientras que fuentes prefrontales aumentadas que parecen reflejar mecanismos de compensación para garantizar una marcha estable.

Palabras clave: trastornos relacionados con el envejecimiento; fuentes electrofisiológicas; MNR modificado; problema inverso; múltiples mínimos cuadrados penalizados; trastorno de la velocidad de la marcha.

INTRODUCTION

Population aging, caused in the first place by the increase in life expectancy and the decrease in the birth rate, is currently one of the most important challenges of society. In these times, older adults live longer, but with an increase in their chronic conditions that can cause functional limitation, disability and dependency, which generates high health care costs, both for the family and for the health system (Abizanda and Rodríguez, 2015). Currently, there is a higher prevalence and a tendency to increase disability in the elderly population, as a consequence of the combination of pathophysiological changes related to aging, chronic diseases and acute or intercurrent processes, all of which are also influenced by the psychosocial environment (Clegg et al., 2013). For this reason, it is necessary to develop tools that allow detecting and differentiating the elderly who present a state of vulnerability and greater risk of developing disability. Vulnerabilities can be reflected both in physical and cognitive performance, which in turn can be a consequence of early structural changes in the brain, suggesting the possibility to find brain biomarkers of physical performance decline with aging. Early evaluation of signs and symptoms related to some functional deterioration would contribute to the early detection of frail elderly people, susceptible to primary and secondary intervention that reduce or delay the appearance of adverse outcomes (Morley, 2016).

Current advances in Neuroimaging techniques have allowed the evaluation of alterations of the central nervous system, in particular the spatial distributions of abnormalities and changes in structural or anatomical brain connectivity. Although these techniques present a very good spatial resolution, they possess low temporal resolution and they are also highly invasive, exposing the subject or patient to radioactive substances or powerful magnetic fields. Besides, the initial and ongoing operating costs along with poor accessibility of Neuroimaging equipment would make impossible a widespread use in older adults. On the other hand, techniques like the Electroencephalography (EEG) offer direct measurements of the electromagnetic activity generated in the brain that is macroscopically reflected on the scalp. These techniques are non-invasive and provide the necessary temporal resolution to study the dynamics of brain processes. Electrophysiological studies have shown high sensitivity for the detection of functional alterations of the brain related to multiple pathologies (Niedermeyer and da Silva, 2005; Moretti et al., 2012; Rodriguez et al., 2014).

In particular, the EEG allows the study of both evoked and spontaneous cerebral electrical activity, whose origin is fundamentally in the cerebral cortex, and could become a research tool to identify those older adults susceptible to disability. However, one of the major drawbacks is the low spatial resolution of the EEG, as it measures electric potentials on a small set of electrodes located on the scalp. To overcoming this issue *in vivo* one needs to solve the so-called EEG inverse problem (IP), which consists in estimating the neural current sources generated inside the brain (primary current density, PCD, **j**), from the electric potentials (voltage differences, **V**) measured with the EEG technique on the scalp surface. The mathematical relation between these magnitudes is given by:

$$\mathbf{V}_{Ne\times 1} = \mathbf{K}_{Ne\times 3Ng} \cdot \mathbf{j}_{3Ng\times 1} + \boldsymbol{\varepsilon}_{Ne\times 1} \qquad (0.1)$$

where the noisy nature of the EEG recordings is explicitly taken into account through the vector of errors ($\boldsymbol{\varepsilon}$). The number of electrodes is Ne and the number of sources Ng corresponds to the number of grid points of the discretization of the source space inside the brain. The $3Ng$ elements of the column vector $\mathbf{j} = \left[j_{1x}, j_{1y}, j_{1z}, \ldots, j_{Ngx}, j_{Ngy}, j_{Ngz}\right]$ correspond to the three components of the PCD vector field for each point in the grid. The matrix $\mathbf{K}$ is known as the Electric Lead Field (ELF), and summarizes the geometric and electric properties of the conducting media (brain, skull and scalp).

Solving this mathematical problem is challenging due to three major problems. First, the number of sources inside the brain is usually much greater than the number of electrodes distributed on the scalp surface ($Ng > Ne$), implying that there exist an infinite number of source spatial configurations that generate the same electric potential distribution on the scalp (non-uniqueness). Second, the system is ill-conditioned because of the high similarity between the columns of the ELF, implying that small changes in the data can lead to totally different configurations in the sources (instability). This is particularly problematic since, in practice, the little errors in measurements due to external (non-brain-related) sources of noise are inevitable. Third, there is no clear way to obtain a ground truth in real EEG experiments, i.e., it is not possible to know, independently of the EEG measurements, what is the underlying spatial distribution of electrical sources in the brain originating these measurements.

These problems are addressed by including additional information on the solution, which can be done by imposing physical or mathematical constraints, or by taking into account anatomical or physiological knowledge about the feasibility of the estimated solutions. However, the additional information selected leads to a specific EEG inverse solution. This, together with the lack of a ground truth, makes difficult to evaluate which method offers a distribution more similar to the real brain activations. Depending on the selected additional information, the large amount of different inverse solutions can be divided into two groups: dipole solutions, which assume that the PCD is a discrete set of current dipoles (Scherg and von Cramon, 1986; Scholz and Schwierz, 1994); and distributed inverse solutions, which assume that the current density is spatially distributed, having a value for each position in the brain (Hämäläinen and Ilmoniemi, 1994; Pascual-Marqui et al., 1994). The dipolar models adjust the orientation and magnitude of an *a priori* fixed number of dipoles, being appropriate when sources are expected to be small but it is not very accurate when estimating spread sources throughout the brain. On the other hand, distributed inverse models are more plausible when estimating sources covering large areas of the brain (Bosch-Bayard et al., 2001; Trujillo-Barreto et al., 2004).

To mathematically addressing the EEG IP, one of the most straightforward and established approach is through its formulation as a penalized linear regression model and the use of Tikhonov regularization (Dale et al., 2000). Two important inverse solutions that follow from this approach are the Minimum Norm (MN) (Hämäläinen and Ilmoniemi, 1994) and the Low-Resolution Electromagnetic Tomography (LORETA) (Pascual-Marqui et al., 1994). However, recent advances in penalized least squares regression have brought attention to the use of non-quadratic penalty functions of the coefficients to estimate, usually based on their l_1 -norm, which can produce sparser solutions (Fan and Li, 2001). It has been shown that they are especially useful when the number of unknowns is much greater than the amount of data (Tibshirani et al., 2005), which is the case in the EEG IP. Some attempts have also been made to propose models with specific combinations of quadratic and non-quadratic penalties (Gorodnitsky and Rao, 1997; Nagarajan et al., 2006). A general multiple penalized least squares (MPLS) model allows us to address combinations of any number of penalty terms, quadratic or not (Valdés-Sosa et al., 2006), and its solution can be found from the optimization problem:

$$\hat{\mathbf{j}} = \underset{\mathbf{j}}{\operatorname{argmin}} \left\{ \mathbf{V} - \mathbf{K}\mathbf{j}^{T} \mathbf{V} - \mathbf{K}\mathbf{j} + \sum_{r=1}^{R} \lambda_r P_r \mathbf{j} \right\} \qquad (0.2)$$

Given the impossibility of knowing which inverse solution is the best one from the many methods proposed (Grech et al., 2008), in this context we follow the strategy of proposing this kind of flexible models that can adjust solutions to the data at hand (Valdés-Sosa et al., 2006). Another advantage of using this model would be the possibility of testing many different solutions in the same framework, which allows a more fair comparison of their performance. Previous works have studied the performance of some particular cases of this general model, such as LASSO, LASSO Fusion, and the Elastic Net, using traditional algorithms in simulated and real EEG data (Vega-Hernández et al., 2008).

In this work, we introduce a general Modified Newton-Raphson (MNR) algorithm to solve any of the models that can be represented by Eq. (1.2), which can be seen as an extension of the Minorization-Maximization approach for these methods. This allows us to compute –within the same computational implementation– several different inverse solutions, as well as addressing novel solutions based on the Adaptive Lasso and Smooth Lasso models that have never been applied to EEG IP before.

METHODOLOGY

MNR ALGORITHM FOR MPLS METHODS

In this section, we present a modified Newton-Raphson (MNR) algorithm to implement Multiple Penalized Least-Squares (MPLS) models. The theoretical background for the formulation of the general MPLS model and a Table of many known and some novel specific models are given in the Appendix (section A). This Appendix also gives a brief literature review on different algorithms that have been proposed/used for estimating specific PLS models.

The MNR algorithm can be seen as an extension of the MM algorithm (Hunter and Li, 2005) for MPLS methods, but are closely related to classical NR techniques. The objective function and its gradient of the MPLS

methods represented by Eq. (1.2) can be re-written in the standard statistical notation (assuming that $g^{r\prime}(0_+) < \infty$, for $r = 1,\ldots,R$) as:

$$f(\boldsymbol{\beta}) = \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \sum_{r=1}^{R}\Psi^r(\boldsymbol{\beta}), \text{ with } \Psi^r(\boldsymbol{\beta}) = \lambda_r\sum_{i=1}^{N_r} g^r(|\theta_i^r|),$$

$$\nabla f(\boldsymbol{\beta}) = -\mathbf{X}^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \sum_{r=1}^{R}\lambda_r\sum_{i=1}^{N_r}\nabla\theta_i^r(\boldsymbol{\beta})\, g^{r\prime}(|\theta_i^r|)\,\operatorname{sgn}(\theta_i^r)$$

where $\theta^r(\boldsymbol{\beta}) = \mathbf{L}^r\boldsymbol{\beta}$ can be seen as a vector function that models the correlation structure of $\boldsymbol{\beta}$. The linear operators $\mathbf{L}^r \in \mathbb{R}^{N_r\times p}$ can be set to $\mathbf{L}^r = \mathbf{I}_p$ to imply independence of β_i or to any other matrix structure (e.g. first and second differences operators that are used to model sparse and smoothness, respectively). The scalar function $\theta_i^r = \mathbf{L}_{i:}^{r\,T}\boldsymbol{\beta}$ represents the i-th entry of $\boldsymbol{\theta}^r(\boldsymbol{\beta})$, where $\mathbf{L}_{i:}^r$ is the i-th row of $\mathbf{L}^r$, while the sign function is represented as "sgn".

We can now follow the same rationale used by the minorization-maximization (MM) algorithm of Hunter and Li (2005). To avoid numerical problems when $\theta \approx 0$, they proposed to use an approximation of the objective function ($f_\varepsilon(\boldsymbol{\beta})$), obtained by perturbing every function g^r for $r = 1,\ldots,R$, using some small $\varepsilon > 0$, as:

$$g_\varepsilon(|\theta|) = g(|\theta|) - \varepsilon\int_0^{|\theta|}\frac{g'(t)}{\varepsilon+t}\,dt$$

The local quadratic approximation of this equation for all penalty functions in the MPLS model leads to:

$$g_\varepsilon^r(|\theta|) \approx g^r(|\theta_i^r|) + \frac{\theta^2 - \theta_i^{r\,2}}{2(\varepsilon+|\theta_i^r|)}\,g^{r\prime}(|\theta_i^r|_+)$$

The NR technique can then be used to locally minimize the perturbed objective function $f_\varepsilon(\boldsymbol{\beta})$ through their first and second derivatives:

$$\begin{aligned}\nabla f_\varepsilon(\boldsymbol{\beta}) &= -\mathbf{X}^T\mathbf{y} + \mathbf{X}^T\mathbf{X} + \sum\lambda_r\mathbf{L}^{r\,T}\mathbf{D}^r\mathbf{L}^r\boldsymbol{\beta},\\ \nabla^2 f_\varepsilon(\boldsymbol{\beta}) &= \mathbf{X}^T\mathbf{X} + \sum\lambda_r\mathbf{L}^{r\,T}\mathbf{D}^r\mathbf{L}^r\end{aligned}$$

where $\mathbf{D}^r = \operatorname{diag}(d_1^r, d_2^r,\ldots,d_{N_r}^r)$ with $d_i^r = g^{r\prime}(|\theta_i^r|)/(\varepsilon+|\theta_i^r|)$ for $i = 1,\ldots,N_r$ and some very small $\varepsilon > 0$. The solution is found for some $\alpha_k > 0$, using the iterative formula:

$$\begin{aligned}\boldsymbol{\beta}^{k+1} &= \boldsymbol{\beta}^k - \alpha_k\left(\nabla^2 f_\varepsilon(\boldsymbol{\beta}^k)\right)^{-1}\nabla f_\varepsilon(\boldsymbol{\beta}^k)\\ &= \boldsymbol{\beta}^k + \alpha_k\left[\mathbf{X}^T\mathbf{X} + \sum\lambda_r\mathbf{L}^{r\,T}\mathbf{D}^r\mathbf{L}^r\right]^{-1}\left[\mathbf{X}^T\mathbf{y} - \boldsymbol{\beta}^k\right]\end{aligned}$$

Note that $\|f(\boldsymbol{\beta}) - f_\varepsilon(\boldsymbol{\beta})\| \to 0$ and $\|\nabla f(\boldsymbol{\beta}) - \nabla f_\varepsilon(\boldsymbol{\beta})\| \to 0$ uniformly whenever $\varepsilon \to 0$. Thus, any limit point of the estimated sequence $\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \ldots$ represents a critical point of the original objective function $f(\boldsymbol{\beta})$ (Hunter and Li, 2005). In our case, the objective function may be more complex if it contains convex and concave penalty functions with correlation structure. In those cases, the MNR algorithm can be stuck at saddle or local stationary points. However, the function $f(\boldsymbol{\beta})$ is convex for the known PLS methods based on the LASSO-type penalty, e.g. FdLASSO (Tibshirani et al., 2005), FnLASSO (Land and Friedman, 1996) and

SLASSO (Hebiri, 2008). Thus, for these cases the MNR implementation and specifically, the canonical version ($\alpha_k = 1$) achieves the global minimum. The parameter ε can be selected similarly to Hunter and Li (2005):

$$\varepsilon = \frac{\tau}{2RM} \min \left\{ \left| \theta_i^r \right| : \theta_i^r \neq 0 \right\}, \text{ for } i = 1,\ldots,\ell_r \text{ and } r = 1,\ldots,R,$$

where $M = \max\{g^{r\prime}(0_+)\}$, for $r = 1,\ldots,R$, and $\tau > 0$ is the convergence parameter $|\partial_j f_\varepsilon(\boldsymbol{\beta})| < \tau/2$. The parameter ε becomes smaller through iterations, but it is usually fixed after the first iteration to avoid numerical instability (Hansen, 1998).

The canonical version of the MNR algorithm extended here for MPLS models is as follows:

MNR algorithm for MPLS ($\mathbf{y} \in \mathbb{R}^{n\times 1}$, $\mathbf{X} \in \mathbb{R}^{n\times p}$, $\lambda_1,\ldots,\lambda_R$, $\mathbf{L}^1,\ldots,\mathbf{L}^R$)

1. Start with $k: k \leftarrow 0$ and set $\tau \leftarrow 10^{-8}$, $\varepsilon \leftarrow 10^{-8}$, $\text{maxiter} \leftarrow 100$ and $\Omega \leftarrow \mathbf{I}_p$.
2. Set $k \leftarrow k+1$ and compute $\boldsymbol{\beta}^k \leftarrow \left(\mathbf{X}^T\mathbf{X} + \Omega\right)^{-1}\mathbf{X}^T\mathbf{y}$.
3. Set $\boldsymbol{\theta}^r \leftarrow \mathbf{L}^r\boldsymbol{\beta}^k$ for $r = 1,\ldots,R$ and compute

   $$\mathbf{D}^r \leftarrow \text{diag}\left(g^{r\prime}\left(\left|\theta_1^r\right| / \varepsilon + \left|\theta_1^r\right|\right),\ldots, g^{r\prime}\left(\left|\theta_{\ell_r}^r\right| / \varepsilon + \left|\theta_{\ell_r}^r\right|\right)\right)$$

4. If $k = 1$, then set $M \leftarrow \max\{g^{r\prime}(0_+)\}$ and $\varepsilon \leftarrow \frac{\tau}{2RM}\min\left\{\left|\theta_i^r\right| : \theta_i^r \neq 0\right\}$.
5. Set $\Omega \leftarrow \sum \lambda_r \mathbf{L}^{r\,T}\mathbf{D}^r\mathbf{L}^r$ and compute $\boldsymbol{\delta} \leftarrow -\mathbf{X}^T\mathbf{y} + \left(\mathbf{X}^T\mathbf{X} + \Omega\right)\boldsymbol{\beta}^k$.
6. If $\left|\delta_j\right| < \tau/2$ for all $j \in 1,\ldots,p$ such that $\left|\beta_j\right| \geq \varepsilon$, then go to Step 8.
7. If $k < \text{maxiter}$, then go to Step 2.
8. Ending step: if convergence is reached then the solution is $\hat{\boldsymbol{\beta}} \leftarrow \boldsymbol{\beta}^k$.

This algorithm depends on regularization parameters $\lambda_1,\ldots,\lambda_R$, which can be selected from a given grid of values or automatically determined according to the singular values of $\mathbf{X}$. The selection of the 'optimal values' for these parameters is a complex process that will not be considered here in detail. This is usually done by evaluating information criteria such as Akaike (AIC), Bayesian (BIC) or generalized cross-validation (GCV). For this purpose, the estimation of the degrees of freedom can be done as proposed in Hunter and Li (2005). In order to avoid the selection of optimal parameters in an R-dimensional grid, in our implementation we prefer to set $\lambda_r = \lambda\mu_r$ and the proportions $\mu_r > 0$, for $r = 1,\ldots,R$, subject to $\sum \mu_r = 1$. The μ_r represent prior assumptions about relative penalty contribution and allow simplifying the selection process to estimate only the parameter λ. In the analysis of real EEG noisy data, it is not easy to ensure that the heuristic criteria would lead to the best trade-off between data fitting and penalization, i.e. to find the optimal value of the regularization parameters. Although there is no way to accurately know the active sources in the brain, in these situations it is always possible to use physiological and empirical knowledge about the specific brain process in study to select the most appropriate representation of its electrophysiological sources, among those given for different regularization parameters.

NOVEL INVERSE SOLUTIONS BASED ON THE ADAPTIVE LASSO

One particular model that can also be handled with the latter approach is the Adaptive LASSO (Zou, 2006), whose objective function is:

$$f\ \boldsymbol{\beta}\ = \frac{1}{2}\ \mathbf{y}-\mathbf{X}\boldsymbol{\beta}\ ^{T}\ \mathbf{y}-\mathbf{X}\boldsymbol{\beta}\ +\lambda\sum_{j}\gamma_{j}\left|\beta_{j}\right|,\quad \gamma_{j}\geq 0\,.$$

This means that the MNR algorithm can be exploited to implement the nonnegative garrote (NNG) method (Breiman, 1995), which is stated as:

$$\hat{\mathbf{w}}=\underset{\mathbf{w}}{\operatorname{argmin}}\left\{\frac{1}{2}\left\|\mathbf{y}-\sum_{j=1}^{p}\mathbf{x}_{j}\beta_{j}^{ols}w_{j}\right\|_{2}^{2}+\lambda\sum_{j=1}^{p}w_{j}\right\};\ \text{with}\ w_{j}=\frac{\beta_{j}}{\beta_{j}^{ols}}\geq 0,$$

where $\boldsymbol{\beta}^{ols}$ represents the OLS estimator. This is analogous to the following formulation:

$$\hat{\boldsymbol{\beta}}=\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left\{\frac{1}{2}\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|_{2}^{2}+\lambda\sum_{j=1}^{p}\frac{1}{\left|\beta_{j}^{ols}\right|}\left|\beta_{j}\right|\right\};\ \ \text{s.t. sgn}(\beta_{j})=\text{sgn}(\beta_{j}^{ols})$$

This can be seen as an extension of the Adaptive LASSO method with sign constraints over $\boldsymbol{\beta}$ and weights $\gamma_{j}=1/\beta_{j}^{ols}$.

As originally conceived, the nonnegative Garrote is limited to $p<n$ situations; basically, because it depends heavily on the OLS estimator. However, it can be extended to the $p\geq n$ scenario while keeping its natural flavor if we consider making it dependent on other reference estimators $\boldsymbol{\beta}^{ref}$ and establish a more general approach (e.g. LASSO, ENET, SLASSO). Based on this model, it is evident that we can propose other extensions to include smoothness constraints into the NNG. A simple representation only uses an additional quadratic term, which we will call the smooth nonnegative Garrote (SNNG), as follows:

$$\hat{\boldsymbol{\beta}}=\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left\{\frac{1}{2}\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|_{2}^{2}+\frac{\lambda_{sm}}{2}\left\|\mathbf{L}\boldsymbol{\beta}\right\|_{2}^{2}+\lambda\sum_{j=1}^{p}\frac{1}{\left|\beta_{j}^{ref}\right|}\left|\beta_{j}\right|\right\};\ \ \text{s.t. sgn}(\beta_{j})=\text{sgn}(\beta_{j}^{ref})$$

which can be shown to be equivalent to:

$$\hat{\mathbf{w}}=\underset{\mathbf{w}}{\operatorname{argmin}}\left\{\frac{1}{2}\left\|\mathbf{y}-\sum_{j=1}^{p}\mathbf{x}_{j}\beta_{j}^{ref}w_{j}\right\|_{2}^{2}+\frac{\lambda_{sm}}{2}\left\|\tilde{\mathbf{L}}\mathbf{w}\right\|_{2}^{2}+\lambda\sum_{j=1}^{p}w_{j}\right\};\ \ \text{s.t. }w_{j}\geq 0$$

where $\tilde{\mathbf{L}}=\mathbf{L}\mathbf{D}$ and $\mathbf{D}=\text{diag}\ \beta_{1}^{ref},\beta_{2}^{ref},\ldots,\beta_{p}^{ref}$. This straightforward extension of the NNG with a smoothness term can also be seen as a type of Smooth LASSO (SLASSO) with weights in the l_{1}-norm term defined by the reference solution and non-negative constraints. Therefore, we also propose to test here the non-negative version of the SLASSO which is equivalent to the SNNG but without the weights of the reference solution, i.e. setting all of them to one. We will call this model as the Non-negative Smooth LASSO (NNSL). In these two novel methods, the use of the proposed MNR algorithm does not directly ensure the nonnegativity of the solution. The use of an active set algorithm, similar to LARS, can provide a natural framework for including sign constraints. However, in practice, we restrict the solutions given by our MNR algorithm to have nonnegative values using the simple procedure of setting to zero all coefficients that take negative values after convergence of the estimation process.

RESULTS AND DISCUSSION

SIMULATION STUDIES

In this section, we use the proposed MNR algorithm for solving the EEG inverse problem with multiple penalties. We firstly simulated a one-dimensional toy example to compare the estimators obtained with the MNR and those obtained using standard algorithms. We also use a similar simulated example to test the SNNG estimator in contrast with other regularization methods. Then, we study a more realistic EEG simulation for a preliminary comparison between different inverse solutions using the MNR algorithm with the equivalent solutions using LARS algorithm, in terms of quality measures (localization error, visibility and blurring).

Comparing MNR to classical procedures in MPLS models

As a first study, we compare the performance of the proposed algorithm in the estimation of models based in the combination of two penalties that have been already calculated by other classical procedures: FdLASSO and SLASSO. We also test the MNR algorithm in the calculation of a model based in a single penalty: FnLASSO. The FdLASSO is currently implemented using an interior point algorithm as shown by Tibshirani et al., (2005), while SLASSO and FnLASSO are implemented using a very well-known transformation to a LASSO problem, which in turn is implemented with the LARS algorithm (Tibshirani et al., 2005; Zou and Hastie, 2005).

In all examples, we simulate $n = 100$ observations and $p = 200$ predictors using the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\beta_j = \exp(-0.015\ j - 50^{\ 2})$ for $30 < j < 70$, $\beta_j = 1$ for $95 \leq j \leq 105$ and $j = 150$, while $\beta_j = 0$ otherwise. The correlation between any pair of predictors $\mathbf{x}_i$ and $\mathbf{x}_j$ is varied with values $\rho =$ 0.5, 0.75, 0.9, 0.95, 0.99; while the components of $\mathbf{x}_j$ and $\boldsymbol{\varepsilon}$ are standard normal. This allows to assess the effect of correlated predictors on the convergence of the MNR procedure.

In the FdLASSO method, we reformulate the problem using the auxiliary parameter $0 < \mu < 1$, i.e., the penalty function is rewritten as $\Psi = \lambda\ \mu g^{\ 1} + 1 - \mu\ g^{\ 2}$, where $g^{\ 1}\ \boldsymbol{\beta} = \|\boldsymbol{\beta}\|_1 = \sum_j |\beta_j|$, $g^{\ 2}\ \boldsymbol{\beta} = \|\mathbf{L}\boldsymbol{\beta}\|_1 = \sum_j |(\mathbf{L}\boldsymbol{\beta})_j|$, with $\mathbf{L}$ being the first differences operator. In this example, we set $\mu = 0.001$ to enhance the smoothness features in the solution and we select the parameter λ that minimizes the BIC criterion evaluated in a predetermined range. The FdLASSO was implemented using the proposed MNR algorithm, such that after selecting the optimal λ parameter, the final estimator $\boldsymbol{\beta}^{MNR}$ was computed. We compare this solution with $\boldsymbol{\beta}^{S}$, which is the one given by the standard procedure (Tibshirani et al., 2005).

A boxplot of the maximum of absolute differences between the MNR and classical estimators, i.e. $\max\{|\beta_j^{MNR} - \beta_j^{S}| : j = 1, \ldots, p\}$, was computed over 50 simulated datasets for each value of the correlation among predictors, as shown in Figure 1 (right panel). As can be seen, the median of the maximum differences are lower than 5% of the maximum coefficient value (which is the unit), for all correlations among predictors. Hence, the coefficients plot for $\boldsymbol{\beta}^{MNR}$ and $\boldsymbol{\beta}^{S}$ are almost superimposed, as shown in Figure 1 (left panel) for a particular dataset with $\rho = 0.5$.

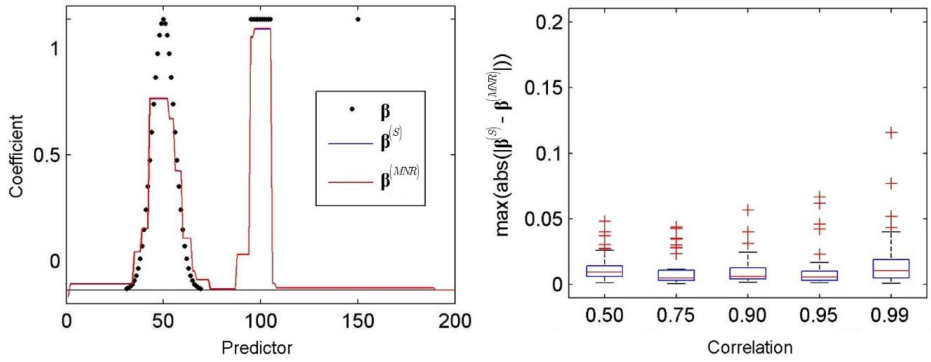

Fig. 1. *Simulated example for comparing FdLASSO solutions implemented with the classical procedure ($\boldsymbol{\beta}^{S}$) and with the MNR algorithm ($\boldsymbol{\beta}^{MNR}$). Left: plot of estimated and simulated coefficients for a dataset with correlation among predictors of $\rho = 0.5$. Right: boxplot of maximum of absolute differences between the MNR and classical estimators for 50 simulated datasets for different values of the correlation between predictors.*

Using the same simulated examples, we computed the boxplot of the maximum of absolute differences between the MNR and classical estimators for the SLASSO and FnLASSO. Figure 2 shows that the MNR algorithm also allows obtaining very accurate estimators for these methods, which are mainly affected when the correlation between predictors is very high (over 0.95). The latter does not have a direct consequence over the convergence of the algorithm; the hidden cause here is that the selected value for the parameter λ did not lead to an effective regularization. In this situation ($p \gg n$), very small or high values for λ lead to severe ill-conditioned problems that are difficult to solve with accuracy with the current techniques. This problem is avoided when the penalty function is appropriate and the regularization parameters are properly selected.

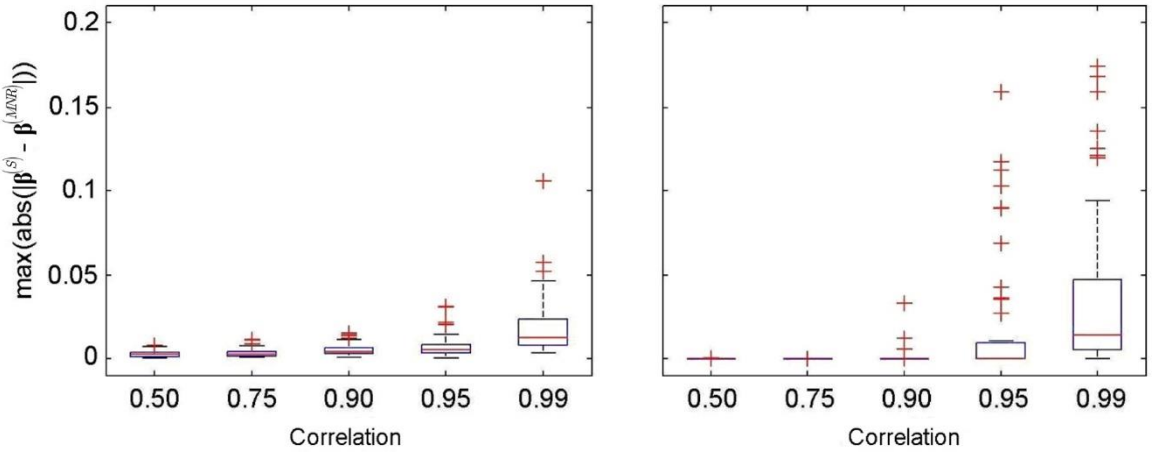

Fig. 2. *Simulated example for comparing FnLASSO (left) and SLASSO (right) solutions implemented with the classical procedure LARS and the MNR algorithm. The boxplot of the maximum of absolute differences between both estimators are shown for each case, considering 50 simulated datasets and different values of correlation among predictors (shown in the abscissa).*

A preliminary study of the newly proposed estimators, the smooth nonnegative garrote (SNNG) and the nonnegative Smooth LASSO (NNSL) is also presented here using the same set of simulated data, comparing their solutions versus Ridge L and SLASSO estimators. Figure 3 shows typical solutions for these methods, as we checked that the effect of correlated predictors was minimal (i.e. estimators from differently correlated predictors were very similar). The SNNG estimator seems to be an improved version of the Ridge L estimator: the most salient features are enhanced and the less significant ones are discarded. In general, the use of the non-negativity constraint in the NNSL, and the use of a previous estimator in the SNNG, allow obtaining sparser solutions. This suggests that these models produce simpler solutions which are more interpretable in statistical analyses.

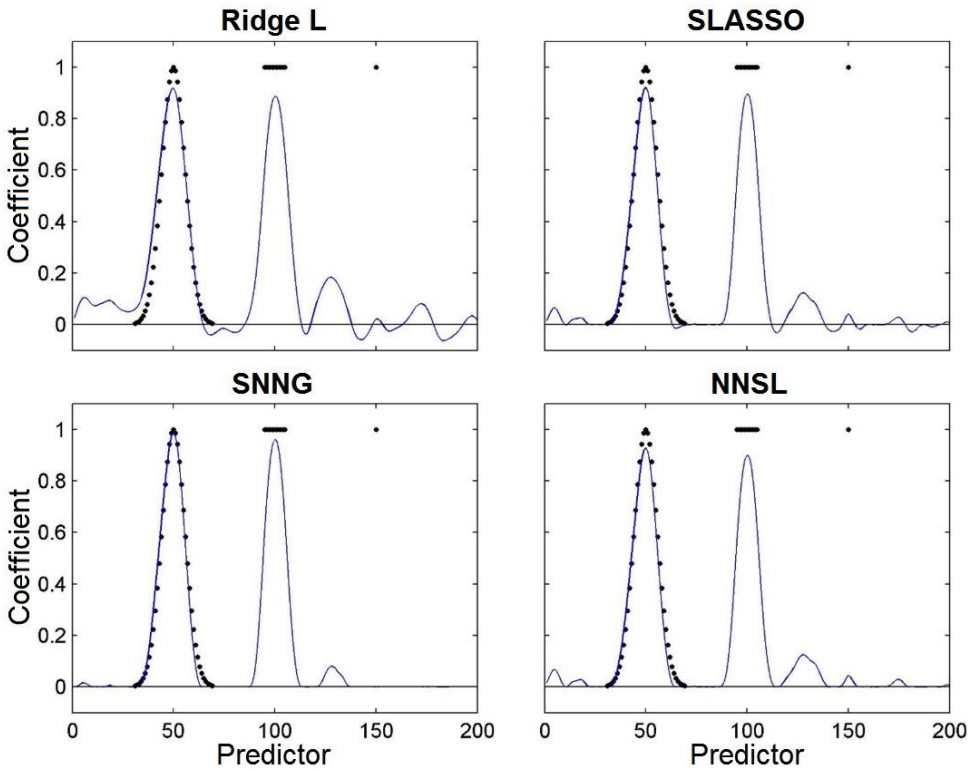

Fig. 3. *Simulated example for comparing solutions given by the Ridge L, Smooth LASSO, SNNG and NNSL models, applied on the same dataset. The last three methods were implemented with the MNR algorithm proposed in this study.*

In all cases, the estimator corresponding to the isolated coefficient suffers from the insufficient data problem, since the use of a smoothness requirement, leads this coefficient to be over-shrunk. In particular, the SNNG method makes a bad estimation of the isolated coefficient because it uses a very high weight for this component. Note that this is very small in the Ridge L estimator. In this simulated example, we do not make preferential distinctions between the SNNG and the NNSL, but we conjecture that if the reference estimator is a good approximation, then the SNNG estimator would be the best alternative.

EEG realistic simulations

The synthetic data used to explore the performance of the MNR algorithm was based in creating two groups of current density distributions. The first group includes three primary current density (PCD) distributions, each one simulated as a spherical three-dimensional Gaussian source with amplitude of 10 nA/mm2 and width of 1 mm (sparse simulation), but with the maximum located in three different anatomical structures of a brain space of 3862 generators: postcentral gyrus (Postcentral), occipital pole left (Occipital) and cingulate region left (Cingulate). The Talairach Coordinates (Talairach and Tournoux, 1988) of the maximum value of each simulated PCD appear in the top row of Figure 4. The second group consisted in three different sets of current density distributions. Each set contained seven simulated primary current densities: a "centroid" PCD with maximum in the same location and anatomical structures of those in the first group, and the others derived from this centroid by moving the maxima in just one grid point. All these PCDs were also simulated as spherical three-dimensional Gaussian sources but with amplitude and width of 10 nA/mm2 and 10 mm (smooth simulation), respectively. "Centroid" PCDs of this second group are also shown in Figure 4.

The simulated voltages $\mathbf{V}$ were obtained through Eq. (1.1), where $\mathbf{j}$ is the simulated PCDs, $\mathbf{K}$ is the Lead Field (computed for an array of 19 electrodes from the 10/20 system, using the common average reference). Additive Gaussian noise was generated with zero mean and standard deviation σ_{noise}, to have a peak signal-to-noise ratio ($pSNR = 20\log_{10} \max V / \sigma_{noise}$) of about 15 db.

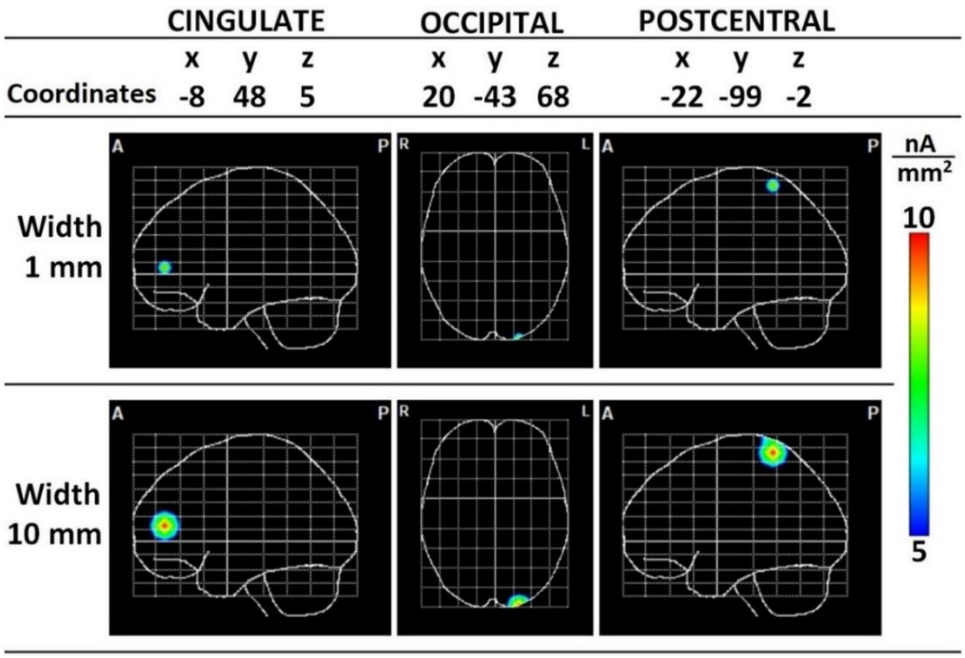

Fig. 4. *Sagittal (Postcentral, Cingulate) and Coronal (Occipital) views of six simulated PCDs. The first group of sources (middle row) were simulated as spherical 3D Gaussian sources with width of 1 mm, while the second group (bottom row) corresponded also to spherical 3D Gaussian sources of 10 mm-width. In both cases, sources had the maxima located in the same places, determined by the Talairach coordinates shown in the first row.*

Figure 5 shows the estimated inverse solutions for the first group of simulations, corresponding to the Ridge model, the LASSO and FnLASSO (both calculated using LARS and MNR) and the ENET L estimated with the MNR algorithm. As expected, the Ridge L solution offered overspread activations with some local maxima near the simulated ones but showing other "ghost" or spurious sources that make this solution difficult to interpret. LASSO and FnLASSO estimators offered sparse sources, but with more difficulties in locating the maxima and showing some ghost sources too. In this example, the models computed with the MNR algorithm were very similar in localization to those computed with the LARS algorithm, but with a slightly increased sparsity. Finally, the ENET L solution is usually smoother than those estimators based on l_1-norm, but still with a similar level of sparsity as the simulated source. Although some of the solutions showed sources near the simulated maxima, in general, very sparse activations are difficult to locate with any of the models, as we found in the previous section.

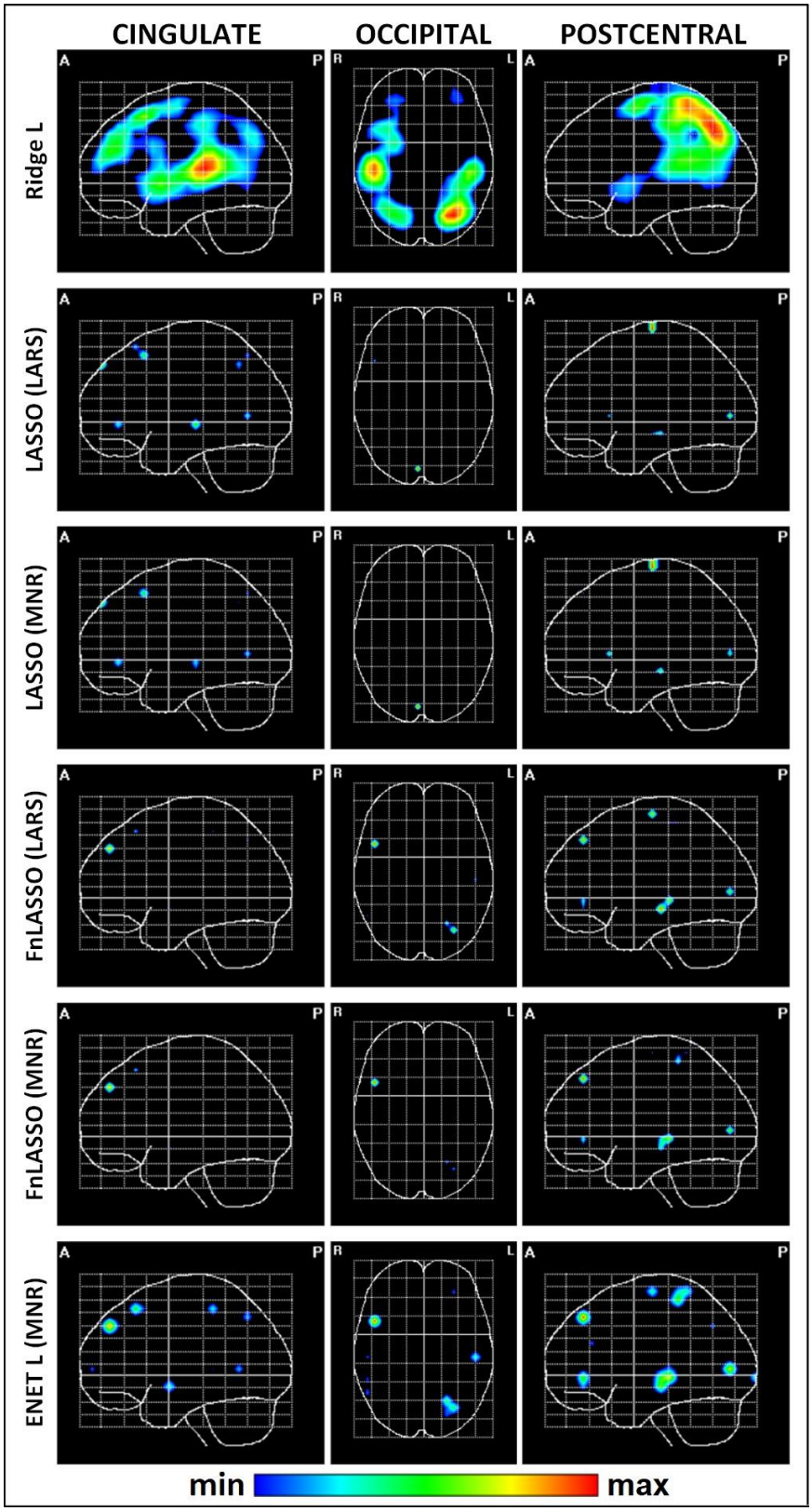

Fig. 5. *Sagittal ((Postcentral, Cingulate) and Coronal (Occipital) views of estimated PCDs for simulations created as spherical 3D Gaussians of 1mm-width. Solutions for the six different models were thresholded at 50% of its maximum value, i.e. the minimum value was set to the half of the maximum value in each case.*

Figure 6 shows the PCDs estimated from the data simulated using the "centroid" PCDs of the second group of non-sparse simulations. Ridge L was estimated using classical Tikhonov regularization and the LQA-Fext algorithm was also used to compute the SLASSO model. In general, all methods offered a much better reconstruction of true activations, showing mostly a single source and with a similar location. LASSO and FnLASSO produced sparse solutions with very similar localization of their maxima and that of the main source of the Ridge L solution, which anyway shows another maximum and a more spread source distribution. The bottom row shows the estimators that imply a combination of l_1-norm and l_2-norm penalties: SLASSO and ENET L. These solutions present sources with localization similar to that obtained from the models using the l_1-norm penalty, but with an intermediate level of sparseness/smoothness with respect to the l_2-norm solution. The level of smoothness varied for the solution estimated with the MNR and the LQA-Fext algorithm (Sánchez-Bornot *et al.*, 2008). In both groups of simulations, the "combination" methods SLASSO and ENET L showed an intermediate level of sparsity between LASSO and Ridge L as they seemed to adjust to the better estimator of the source, so they appear to be a better alternative when ground truth is not known.

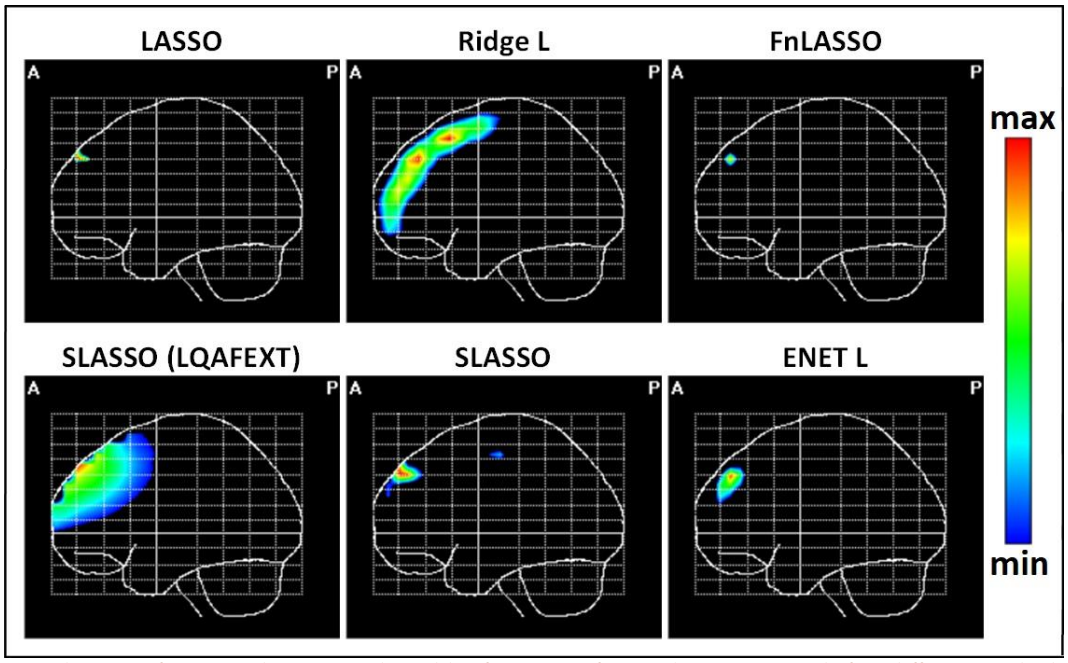

Fig. 6. *Sagittal views of estimated PCDs with width of 10 mm of Cingulate region with five different methods: LASSO, Ridge L, FnLASSO, SLASSO and ENET L using the MNR algorithm and the SLASSO estimator using the LQA-Fext algorithm (bottom left). In every case the minimum was set to 50% of the maximum value.*

Using the whole set of simulations in the second group, we compared the solutions of the different methods in terms of localization error, visibility and blurring. These are well-known measures of the quality of the reconstruction that assess the ability to correctly locate the maximum, estimate its amplitude and estimate its width, respectively (Pascual-Marqui, 1999). Appendix (section B) shows their mathematical definitions together with an algebraic transformation of each measure to take them to a normalized range, such that perfect reconstructions lead to all measures being equal to 1. Based on the information of the quality measures computed for each inverse solution across the 7 locations in each simulation set, we created Figure 7, which summarizes the general tendency of these results and illustrates it with the centroid solution for the simulated Postcentral activation.

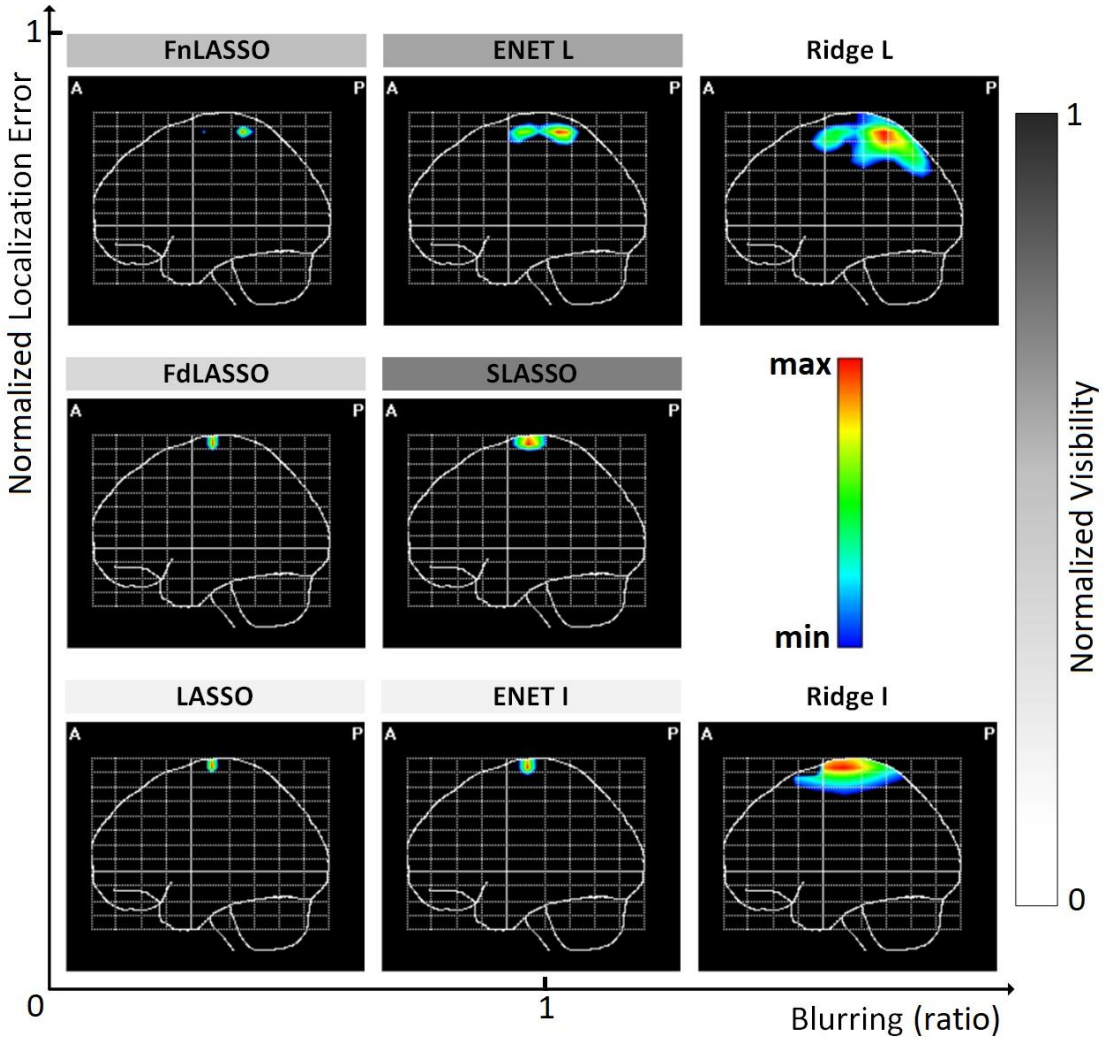

Fig. 7. *Summary of comparison of different inverse solutions estimated by MNR algorithm in terms of the quality measures (normalized localization error (y-axis) vs. non normalized blurring (x-axis), while the normalized visibility is coded in a grayscale which is used as the background color of the name of each solution). The qualitative position of each model in this space is illustrated with the PCDs estimated from the Postcentral simulated centroid source (10 mm, see Figure 4 bottom-right panel), thresholded to 50% of the maximum value.*

As the simulated source has a 10 mm-width, we expected that the models based on l_1-norm (left column, LASSO, FdLASSO and FnLASSO) led to over-sparse solutions, reflected in the lowest blurring, while Ridge solutions based on l_2-norm were usually over-smooth, leading to blurring higher than 1. Those solutions based on the combination of l_1-norm and l_2-norm penalties (ENET and SLASSO) were able to reconstruct the activation with widths more similar to the real one, giving the values of blurring closer to 1. It is interesting to see that the visibility and blurring are not completely independent measures. The better the estimation of the blurring, the better the visibility. This could be explained by the re-distribution of the energy in the estimated solution, such that very sparse solutions concentrate the energy in a few voxels having values higher than the true simulated ones (i.e. visibility higher than 1, normalized visibility far from ideal value of 1), while the very smooth solutions distributes the energy in many nonzero activations, making the maximum to be smaller than the simulated one (i.e. very low visibility).

Another interesting result is that using the Laplacian operator seemed to help find the correct location of the maximum activation, as the versions that did not use this operator in the penalty term (LASSO, ENET I and Ridge I) located the maxima to a more superficial location, closer to the sensors. The ENETL, SLASSO and FnLASSO showed the best localization and smaller blurring than Ridge L. This is in accordance with previous studies on linear solutions that claim that the Laplacian operator impose a more realistic, higher penalization to voxels in the boundary, as these have less neighbors (Pascual-Marqui, 1999). Although this typically takes the

maximum value to nearby voxels that are indeed closer to the simulated source in many simulations, it can also move the maximum to voxels surrounded by many neighbors, which can be far from the real source if the distance between neighbor voxels is large and/or there are few locations with all neighbor voxels being valid sources.

Finally, it can be seen that SLASSO offered a very neat source, similar to the simulated one, although not perfectly located, while ENET L showed the maximum with a better location but also showing a second ghost activation. This suggests that these models can indeed be more flexible and adapt to the data in order to find sources with intermediate level of sparseness. However, these solutions are based on combination of penalties and therefore they strongly depend on the balance of the different regularization parameters. If the parameters are not well balanced, these solutions can be very similar to one of the extreme cases showing over-sparse or over-smooth sources. Future work is needed to test different options for finding the optimal regularization parameters, although in practice, other statistical, clinical or physiological information might be used to select the solution with the best balance, i.e. leading to sources with the most appropriate width (Vega-Hernandez et al., 2008).

Source imaging of resting-state EEG data in elderly subjects

Tests of physical performance provide objective measurements of the functional state of the elderly, and they have been shown to be effective to identify subjects more susceptible to disability, hospitalization and death (García-Agustín et al., 2018). Walking speed (or gait speed, GS) is a physical measure widely used in the clinical practice as a major predictor to evaluate the physical condition of elderly individuals. A decreasing gait speed has been linked to the global cognitive capacity, the executive function and the memory in non-demented elder adults (García-Agustín et al., 2020). Recently, there has been an increase in the number of studies on gait speed not only as an indicator of health status but also as a predictor of adverse health conditions related to a decline in the brain function in elder adults (Varma et al., 2016; Pinter et al., 2017). Main findings suggest that there is increased brain activation in the prefrontal cortex in response to cognitive tasks during the gait. Given the complex cognitive processes involved in gait speed, it has been hypothesized that the reduction of the motor functions could be an early and sensitive indicator of sub-clinical deficits among cognitively normal individuals (Mielke et al., 2012).

In this work, we attempt to find and compare the electrophysiological sources of the resting-state brain rhythms in elder adults. The data set belongs to a prospective cohort study of the gait speed, performed in 2010 and 2016, involving 96 functionally or physically active, community-dwelling elderly adults over 60 years old that regularly practice mild-to-moderate exercise in their communities. The study is explained in detail in García-Agustín et al. (2020), and it aims at assessing the relationship of disorders in the physical performance with brain diseases, in order to reveal the central mechanism of degeneration. Here we explore the sources of the resting-state EEG measured in the second appointment (in 2016) to evaluate their changes according to the physical performance of the subjects, specifically to the gait speed. For this subset, we selected a balanced sample classified into two groups, depending on the evolution of the gait speed (GS) value registered in 2010 and 2016 evaluations. GS was quantified from measuring the time spent to cover 4 meters at a normal pace and was expressed in meters per second (m/s). The groups had 15 subjects each that were matched in age and defined as follows: Group GG) Subjects that presented a good walk speed in 2010 and preserved their normal speed above 0.8 m/s in 2016. Group BB) Subjects that presented an abnormal speed equal or below 0.8 m/s, both in 2010 and 2016 evaluations. The EEG data were recorded with Neuronic EEG recording system MEDICID 5. Nineteen Ag/AgCl electrodes were placed on the scalp according to the international 10/20 electrode placement system with the reference electrode on linked earlobes. Electrode impedance was kept below 10 kOhm. The application of the Fast Fourier Transform to clean segments of EEG provided the complex amplitude for oscillatory activities from 0.39 Hz to 19.89 Hz (frequency step of 0.39 Hz). We selected the average of those amplitudes in the theta band (4-7 Hz) as the input of the algorithm in order to find the electrophysiological sources of the theta rhythm (Niedermeyer and da Silva, 2005).

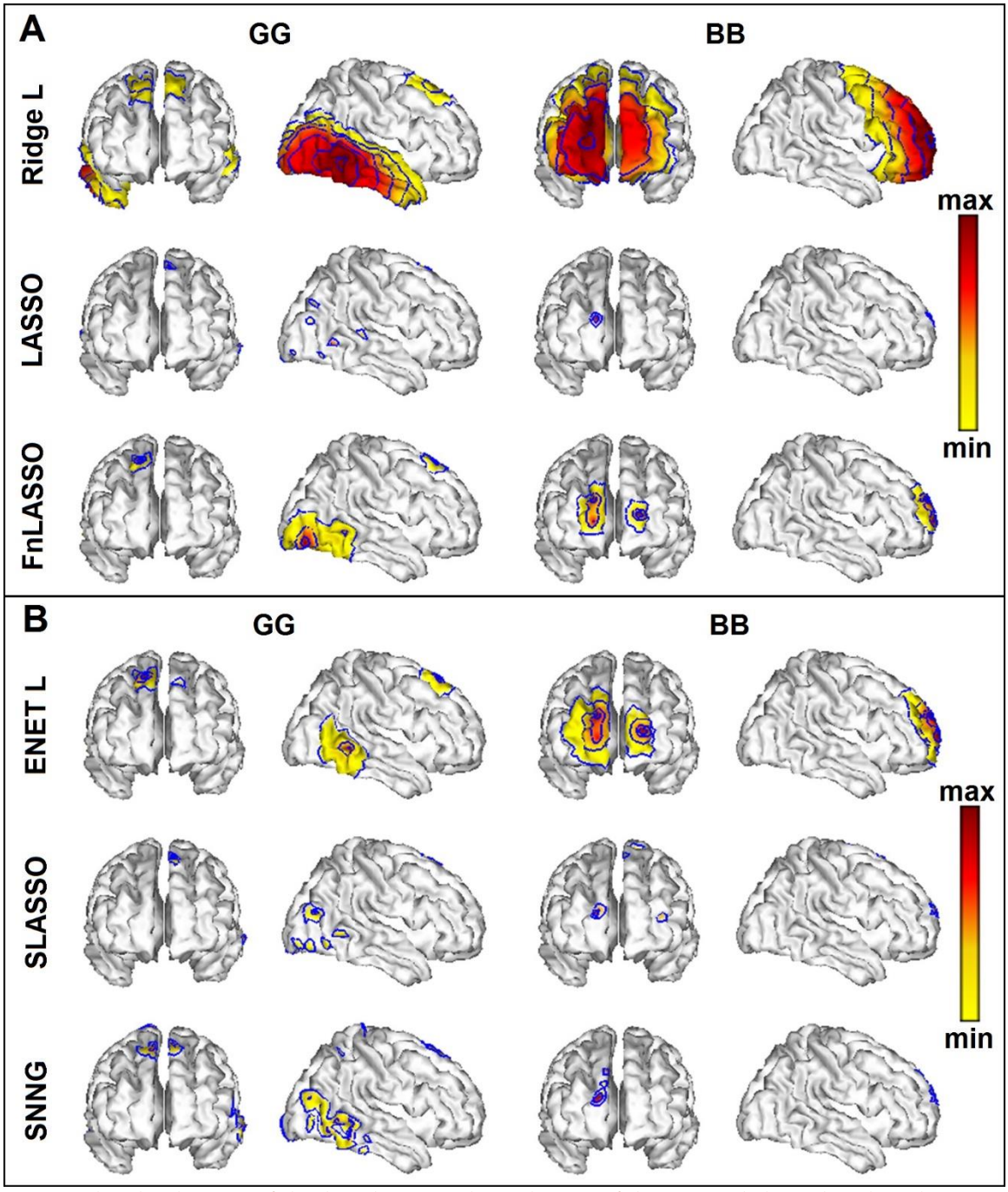

Fig. 8. *Frontal and right views of the three-dimensional visualization of the estimated electrophysiological sources in the cortical surface, averaged across subjects of the two groups of elder subjects (GG: two columns to the left and BB: two columns to the right). Panel **A** shows the solutions based on only one penalty Ridge L, LASSO, LASSO Fusion, while Panel **B** shows the inverse solutions based on the combination of sparseness and smoothness ENET L, Smooth LASSO and Smooth non-negative Garrote. All solutions are color-coded as shown in the colorbar, where the minimum value was taken as the 60% of the maximum in each case. Blue lines represent the iso-contours.*

The different methods showed the expected behavior as studied with simulations: Ridge L offered spread activations, while LASSO offered a very sparse solution. FnLASSO and ENET L showed sources with an intermediate size and SLASSO and SNNG in this case showed sparser solutions than ENET L, but all of them seem more realistic according to the known physiology than those of Ridge L and LASSO. All solutions gave the main activations in similar brain areas although with the maximum source in varied locations.

The maximum of the solutions were located in the bilateral inferior frontal areas (mainly right hemisphere) for BB group and in inferior temporal regions (mainly right) for the GG group. These results are in agreement with

previous studies using only scalp EEG that related the lateralization of frontal theta rhythms with bad performance in gait speed. In a study of 15 elder adults, authors reported asymmetries in slow activities of the alpha and theta bands in subjects with the worst performance in gait speed. In particular, with a model of frontal asymmetry, they reported a significant association between asymmetry in the theta band focused on frontotemporal regions and the state of gait involvement in older adults (Vogt et al., 2010).

Moreover, the different sources found in the two groups might reflect the cerebral mechanisms underlying the mobility which is related to physical performance. The prefrontal cortex is related not only to complex motor functions such as walking, but also to active memory that stores information related to the location of objects in space to guide movement and executive functions that are closely associated with starting actions (start of the movement) (Suzuki et al., 2008). These motor areas connect with each other through cortico-cortical connections to develop the movement strategy and from there connect to the primary motor areas that execute the movement (Varma et al., 2016). Therefore, our results support the hypothesis of Allali and colleagues, who reported that the increased activity of the prefrontal and supplementary cortex observed in elder adults is a reflection of compensation strategies that ensure stability in the rhythm of walking and that fail in those with gait disorders (Allali et al., 2014).

CONCLUSIONS

The main objective of this study is to evaluate the ability of the MNR algorithm to estimate different inverse solutions with several penalizations or constraints. Although the algorithm has been used before in specific studies (Valdés-Sosa et al., 2005), here we formalize it as a general technique and illustrate its potential application for solving MPLS methods. Known techniques such as the LQA and MM algorithms can be seen as variants of the MNR technique but have been constrained to implement particular methods. Using simulated data, we showed that the MNR technique offered similar solutions as other state-of-the-art methods and could be applied for estimating general MPLS models that allows recovering sparse and smooth estimates using the corresponding penalty functions and combinations in a consistent and robust way. Sparsity-inducing models such as LASSO and LASSO Fusion always showed inverse solutions that were focused and concentrated around a few nonzero points. The use of models that combine sparseness and smoothness such as Elastic Net and Smooth LASSO provided solutions with a smooth patches of activations, indeed showing sources with widths that were closer to the size of the real simulated source. The flexibility of the algorithm and its implementation allowed us to explore novel inverse solutions that can include other constraints. In particular, we proposed an extension for the NNG method for the $p>>n$ scenario, which can convey the advantages of that model to this challenging scenario. We showed that the MNR technique can also be applied to general nonlinear optimization problems such as Smooth LASSO or Adaptive LASSO. Combining these models, we proposed a novel solution called Smooth Non-negative Garrote, which can use different reference solutions (besides Ordinary Least Squares) and combine sparsity, smoothness and nonnegativity constraints. We also propose a similar model which do not use a reference solution and therefore it is just a non-negative version of the Smooth LASSO. Future works could establish a general algorithmic framework for exploring other new methods such as extending SCAD (Smoothly Clipped Absolute Deviation) (Fan and Li, 2001) for estimating smooth features.

In this work, we also explored the usefulness of the estimators obtained with the MNR algorithm when applied to the EEG inverse problem. Firstly, we found using realistic simulations that EEG sources could be reconstructed both in the case of sparse, smooth and patch-smooth situations. Solutions showed ghost sources and were not always able to exactly localize the main sources, as in other linear methods (such as LORETA or Ridge L, and Minimum Norm or Ridge I). However, results were consistent with the numerical simulations and in many cases, the additional physiological information about the experiment or the clinical situation could help in extracting relevant information.

Finally, sources of real resting state EEG data from a study of EEG changes related to physical performance in elderly subjects, were obtained with the MNR using both classical and novel models. Results were consistent with the literature about the different sources of theta activity in subjects with preserved or declined physical performance, as reflected by the gait speed. Specifically, results indicate that the increase in the magnitude of theta activity sources in the supplementary and prefrontal areas in elder adults with reduced gait speed (BB group) is related to the deficit in multisensory co-activation in these areas, which function as compensatory mechanisms against peripheral sensory deficit. This suggests that impaired executive functions lead to chronic impairments in gait speed. Altogether, these results serve as a preliminary validation of the MNR methodology,

while at the same time offered a promising hypothesis about the feasibility of using EEG source imaging as potential biomarkers for early physical decline in elder adults.

Several limitations of the MNR algorithm should be studied and addressed in future works. First, similarly to LQA and MM algorithm, the method still gives low nonzero values to coefficients that should be zero and shows a slow convergence. In this sense, we should mention that computing flexible EEG inverse solutions from models that use combination of penalties always takes more time than direct computations such as Ridge (LORETA and Minimum Norm solutions) and other more efficient algorithms such as LARS. Although this can hinder its use in clinical applications, in our study we found that the slowest solution took less than 6 seconds of computational time. Therefore, this would be a relevant limitation only in studies involving a large number of subjects or those needing to compute solutions for many time or frequency points. Second, an important issue is the appropriate choices of regularization parameters with respect to variable selection. The accurate estimation and accurate variable selection can be conflicting goals. Third, the use of nonnegative constraints for the new models proposed here is not optimal. Although some of these issues have been already addressed with other approaches (e.g. a Bayesian approach to learn the regularization parameters), we think that all of these limitations might be effectively overcome with the use of techniques based on selection of variables. These methods perform the selection and estimation processes simultaneously, being the LARS algorithm a good example. They replace the search for an optimal regularization parameter by the stopping criteria in the path of solutions and allow naturally the estimation of only those coefficients that contribute to the solution with a positive (or negative) sign. Therefore, future research should be devoted to the development of an active-set algorithm that can estimate general MPLS models and include nonnegative solutions. Finally, future works should be dedicated also to a more thorough validation of the MNR using other general measures of the quality of the reconstruction such as ROC measures, as well as studying the robustness of the solutions to noise in the realistic EEG setting.

ACKNOWLEDGEMENTS

This work was supported by the VLIR-UOS project "A Cuban National School of Neurotechnology for Cognitive Aging" and the National Fund for Science and Innovation of Cuba. CU2017TEA436A103. Partial funding was also provided by the National Program for Neuroscience and Neurotechnology, Project 15; and from the FONCI mechanism for funding science of the Cuban Ministry of Science, Technology and Environment.

BIBLIOGRAPHIC REFERENCES

Abizanda, P., & Rodríguez, L. (2015). Evaluación funcional. En I. P. Bradley, J (Ed.), Tratado de medicina geriátrica: fundamentos de la atención sanitaria a los mayores (pp. 222-250). Barcelona: Elsevier.

Allali, G., Van Der Meulen, M., Beauchet, O., Rieger, S. W., Vuilleumier, P., & Assal, F. (2014). The neural basis of age-related changes in motor imagery of gait: an fMRI study. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences, 69(11), 1389-1398. https://doi.org/10.1093/gerona/glt207

Bosch-Bayard, J., Valdés-Sosa, P., Virues-Alba, T., Aubert-Vázquez, E., John, E. R., Harmony, T., ... & Trujillo-Barreto, N. (2001). 3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA). Clinical Electroencephalography, 32(2), 47-61. https://doi.org/10.1177/155005940103200203

Breiman, L. (1995). Better subset regression using the nonnegative garrote. Technometrics, 37(4), 373-384. https://doi.org/10.1080/00401706.1995.10484371

Cao, Y., Huang, J., Jiao, Y., & Liu, Y. (2017). A lower bound based smoothed Quasi-Newton algorithm for group bridge penalized regression. Communications in Statistics – Simulation and Computation, 46(6), 4694-4707. https://doi.org/10.1080/03610918.2015.1129409

Chang, L., Roberts, S., & Welsh, A. (2018). Robust lasso regression using Tukey's biweight criterion. Technometrics, 60(1), 36-47. https://doi.org/10.1080/00401706.2017.1305299

Chen, C., ShanXiong, C., Chen, L., & YuChen, Z. (2017). Method for solving LASSO Problem based on multidimensional weight. Advances in Artificial Intelligence, 2017. https://doi.org/10.1155/2017/1736389

Clegg, A., Young, J., Iliffe, S., Rikkert, M. O., & Rockwood, K. (2013). Frailty in elderly people. The lancet, 381(9868), 752-762. https://doi.org/10.1016/S0140-6736(12)62167-9

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron, 26(1), 55-67. https://doi.org/10.1016/S0896-6273(00)81138-1

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. The Annals of statistics, 32(2), 407-499. https://doi.org/10.1214/009053604000000067

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), 1348-1360. https://doi.org/10.1198/016214501753382273

Friedman, J., Hastie, T., Höfling, H., & Tibshirani, R. (2007). Pathwise coordinate optimization. The annals of applied statistics, 1(2), 302-332. https://doi.org/10.1214/07-AOAS131

Fu, W. J. (1998). Penalized regressions: the bridge versus the lasso. Journal of computational and graphical statistics, 7(3), 397-416. https://doi.org/10.1080/10618600.1998.10474784

Fuchs, M., Wagner, M., Köhler, T., & Wischmann, H. A. (1999). Linear and nonlinear current density reconstructions. Journal of clinical Neurophysiology, 16(3), 267-295. https://doi.org/10.1097/00004691-199905000-00006

García-Agustin, D. (2018). Desempeño físico en adulto mayor y su relación con el desenlace y la función cerebral. PhD Thesis, National Center for Scientific Research, Havana, Cuba.

García-Agustin, D., Morgade-Fonte, R. M., Bobes, M. A., Galán-García, L., & Rodríguez-Rodríguez, V. (2020). Association between gait speed decline and EEG abnormalities in a cohort of active older adults living in the community. medRxiv, preprint doi: https://doi.org/10.1101/2020.05.03.20089540.

García-Agustin, D., Soler-Morejon, C. D., & Rodriguez-Perez, Z. (2018). Physical performance tests in the prognosis of adverse outcomes in the elderly. MediSan, 22(06), 466-470.

Gorodnitsky, I. F., & Rao, B. D. (1997). Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. IEEE Transactions on signal processing, 45(3), 600-616. https://doi.org/10.1109/78.558475

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., ... & Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. Journal of neuroengineering and rehabilitation, 5(1), 1-33. https://doi.org/10.1186/1743-0003-5-25

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. Medical & biological engineering & computing, 32(1), 35-42. https://doi.org/10.1007/BF02512476

Hansen, P. C. (1998). Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion. Society for Industrial and Applied Mathematics.

Hebiri, M. (2008). Regularization with the smooth-lasso procedure. arXiv preprint arXiv:0803.0668. https://doi.org/10.48550/arXiv.0803.0668

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67. https://doi.org/10.1080/00401706.1970.10488634

Hunter, D. R., & Li, R. (2005). Variable selection using MM algorithms. Annals of statistics, 33(4), 1617. https://doi.org/10.1214/009053605000000200

Land, S., & Friedman, J. (1996). Variable fusion: a new method of adaptive signal regression. Technical Report, Department of Statistics, Stanford University, Stanford.

Li, R., Dziak, J., & Ma, Y. (2006). Nonconvex penalized least squares: characterizations, algorithm and application. Manuscript.

Mielke, M. M., Roberts, R. O., Savica, R., Cha, R., Drubach, D. I., Christianson, T., ... & Petersen, R. C. (2013). Assessing the temporal relationship between cognition and gait: slow gait predicts cognitive decline in the Mayo Clinic Study of Aging. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences, 68(8), 929-937. https://doi.org/10.1093/gerona/gls256

Moretti, D. V., Zanetti, O., Binetti, G., & Frisoni, G. B. (2012). Quantitative EEG markers in mild cognitive impairment: degenerative versus vascular brain impairment. International Journal of Alzheimer's disease, 2012. https://doi.org/10.1155/2012/917537

Morley, J. E. (2016). Frailty and sarcopenia: the new geriatric giants. Revista de investigacion clinica, 68(2), 59-67.

Mørup, M., Madsen, K. H., & Hansen, L. K. (2008). Approximate ℓ0 constrained non-negative matrix and tensor factorization. In 2008 IEEE International Symposium on Circuits and Systems (pp. 1328-1331). https://doi.org/10.1109/ISCAS.2008.4541671

Nagarajan, S. S., Portniaguine, O., Hwang, D., Johnson, C., & Sekihara, K. (2006). Controlled support MEG imaging. NeuroImage, 33(3), 878-885. https://doi.org/10.1016/j.neuroimage.2006.07.023

Niedermeyer, E., & da Silva, F. L. (Eds.). (2005). Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins.

Pascual-Marqui, R. D. (1999). Review of methods for solving the EEG inverse problem. International journal of bioelectromagnetism, 1(1), 75-86.

Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. International Journal of psychophysiology, 18(1), 49-65. https://doi.org/10.1016/0167-8760(84)90014-X

Pinter, D., Ritchie, S. J., Doubal, F., Gattringer, T., Morris, Z., Bastin, M. E., ... & Wardlaw, J. (2017). Impact of small vessel disease in the brain on gait and balance. Scientific reports, 7(1), 1-8. https://doi.org/10.1038/srep41637

Riera, J. J., & Fuentes, M. E. (1998). Electric lead field for piecewise homogeneous volume conductor model of the head. IEEE transactions on biomedical engineering, 45(6), 746-753. https://doi.org/10.1109/10.678609

Rodriguez, R., Lopera, F., Alvarez, A., Fernandez, Y., Galan, L., Quiroz, Y., & Bobes, M. A. (2014). Spectral analysis of EEG in familial Alzheimer's disease with E280A presenilin-1 mutation gene. International Journal of Alzheimer's Disease, 2014. https://doi.org/10.1155/2014/180741

Sánchez-Bornot, J. M., Martínez-Montes, E., Lage-Castellanos, A., Vega-Hernández, M., & Valdés-Sosa, P. A. (2008). Uncovering sparse brain effective connectivity: A voxel-based approach using penalized regression. Statistica Sinica, 18(4), 1501-1518. https://www.jstor.org/stable/24308566

Scherg, M., & Von Cramon, D. (1986). Evoked dipole source potentials of the human auditory cortex. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 65(5), 344-360. https://doi.org/10.1016/0168-5597(86)90014-6

Scholz, B., & Schwierz, G. (1994). Probability-based current dipole localization from biomagnetic fields. IEEE transactions on biomedical engineering, 41(8), 735-742. https://doi.org/10.1109/10.310089

Suzuki, M., Miyai, I., Ono, T., & Kubota, K. (2008). Activities in the frontal cortex and gait performance are modulated by preparation. An fNIRS study. NeuroImage, 39(2), 600-607. https://doi.org/10.1016/j.neuroimage.2007.08.044

Talairach, J. and Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain-3-dimensional proportional system. An approach to cerebral imaging. New York: Theme Medical Publishers.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(1), 91-108. https://doi.org/10.1111/j.1467-9868.2005.00490.x

Trujillo-Barreto, N. J., Aubert-Vázquez, E., & Valdés-Sosa, P. A. (2004). Bayesian model averaging in EEG/MEG imaging. NeuroImage, 21(4), 1300-1319. https://doi.org/10.1016/j.neuroimage.2003.11.008

Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., & Canales-Rodríguez, E. (2005). Estimating brain functional connectivity with sparse multivariate autoregression. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1457), 969-981. https://doi.org/10.1098/rstb.2005.1654

Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Vega-Hernández, M., Melie-García, L., Lage-Castellanos, A., & Canales-Rodríguez, E. (2006). 18 Granger Causality on Spatial Manifolds: applications to Neuroimaging. In B. Schelter, M. Winterhalder, and J. Timmer (Eds.), Handbook of Time Series Analysis: Recent Theoretical Developments and Applications (pp. 461-941). Freiburg: Wiley-VCH.

Varma, V. R., Hausdorff, J. M., Studenski, S. A., Rosano, C., Camicioli, R., Alexander, N. B., … & Carlson, M. C. (2016). Aging, the central nervous system, and mobility in older adults: interventions. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences, 71(11), 1451-1458. https://doi.org/10.1093/gerona/glw080

Vega-Hernández, M., Martínez-Montes, E., Sánchez-Bornot, J. M., Lage-Castellanos A. and Valdés-Sosa, P.A. (2008). Penalized least squares methods for solving the EEG inverse problem. Statistica Sininica, 18, 1535–1551. https://www.jstor.org/stable/24308568

Vogt, T., Schneider, S., Brummer, V., & Struder, H. K. (2010). Frontal EEG asymmetry: the effects of sustained walking in the elderly. Neuroscience Letters, 485(2), 134-137. https://doi.org/10.1016/j.neulet.2010.09.001

Voronin, S. (2012). Regularization of linear systems with sparsity constraints with applications to large scale inverse problems. PhD Thesis. Princeton University, Princeton, USA.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418–1429. https://doi.org/10.1198/016214506000000735

CONTRIBUCIÓN DE LOS AUTORES:

Mayrim Vega-Hernández: curación de datos, análisis formal, investigación, metodología, software, validación, visualización, redacción del borrador original.

Darío Palmero-Ledón: análisis formal, investigación, validación, visualización, redacción del borrador original.

José M. Sánchez-Bornot: conceptualización, metodología, software, validación.

Jhoanna Perez-Hidalgo-Gato: curación de datos, análisis formal, visualización.

Daysi García-Agustin: adquisición y curación de datos.

Pedro A. Valdés-Sosa: conceptualización, supervisión.

Eduardo Martínez-Montes: investigación, metodología, supervisión, redacción (revisión y edición).

Appendix

Section A. Theoretical Background

Consider a multiple lineal regression model and rewrite Eq. (1.1) using a more standard statistical notation and terminology, as follows:

$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$$

where the columns $(\mathbf{x}_1,\ldots,\mathbf{x}_p \in \mathbb{R}^n)$ of the design matrix $\mathbf{X}$ are the predictors, $\mathbf{y} \in \mathbb{R}^n$ is the response vector, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of coefficients to be estimated and $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is the error term, with the typical simple assumption that $\boldsymbol{\varepsilon} \sim N(\vec{0}, \sigma^2\mathbf{I}_n)$, where σ^2 is the variance of the noise component and $\mathbf{I}_n$ represents the size-n identity matrix. The case $p \gg n$, corresponding to an underdetermined scenario, has led to a huge amount of scientific work on how to estimate models with many new extensions of regularization techniques in the last decade (Voronin, 2012). These techniques produce biased but stable linear solutions when using l_2-norm penalties, being ridge regression (Hoerl and Kennard, 1970) the classical example. The advent of the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996) and the emergence of the more general penalized least squares (PLS) formulation (Fan and Li, 2001), allowed the recovery of sparse solutions, where only a small number of coefficients are nonzero, while Ridge does not produce sparse estimators.

The PLS methods were firstly designed to deal with sparse estimators and thus, they can be seen as variable selection techniques. However, they can also be used to represent methods that lead to smooth estimators and even include combination of penalty terms that may involve concave and convex functions. In this broad sense, the Fused LASSO (FdLASSO) (Tibshirani et al., 2005), together with the Elastic Net (ENET) (Zou and Hastie, 2005) and the Smooth LASSO (SLASSO) (Hebiri, 2008), can be seen as particular instances. In general, these PLS methods can be represented as instances of a more general model that has been named as Multiple PLS (MPLS) (Vega-Hernández et al., 2008; Sánchez-Bornot et al., 2008), and can be expressed as:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left\{\frac{1}{2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^T\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right) + \Psi\left(\boldsymbol{\beta}\right)\right\} \qquad \text{(A.1)}$$

where the functional term $\Psi\left(\boldsymbol{\beta}\right) = \sum_{r=1}^{R}\lambda_r\sum_{i=1}^{N_r} g^{(r)}\left(\left|\theta_i^{(r)}\right|\right)$ represents a sum of several convex and non-convex constraints or penalty terms. The penalty functions $g^{(r)}: \mathbb{R} \mapsto \mathbb{R}$, for $r = 1,\ldots,R,$ are symmetric, non-negative, non-decreasing and continuous over $0,+\infty$. They are in general evaluated at the components of vectors $\boldsymbol{\theta}^r = \mathbf{L}^r\boldsymbol{\beta}$, where $\mathbf{L}^r \in \mathbb{R}^{N_r\times p}$ are linear operators, e.g., the matrix of first or second differences. The regularization parameters λ_r establish the relative importance of the corresponding constraints. It can be easily shown that LASSO and Ridge are particular instances of this equation by setting $R = 1$, $\mathbf{L} = \mathbf{I}_p$ and using the corresponding l_1 and l_2 penalty functions. These and other examples are given in Table 1.

On the other hand, there are a few variable selection procedures that cannot be represented as a PLS method, such as the "best subset selection". For coping with the well-known instability of this method, Breiman introduced the Nonnegative Garrote (NNG) as a variable selection technique that shrinks and zeroes the ordinary least squares (OLS) estimator in order to give intermediate results between OLS and subset selection (Breiman, 1995). However, one important limitation of NNG is that is restricted to the $p < n$ case, which hinders its applications to general problems. In this work we notice that the NNG can indeed be included in the general MPLS formulation, as shown in Table 1.

Table 1: Known models represented as instances of equation A.1. Here $\mathbf{L}$ is the second differences operator and $\mathbf{\Omega}$ can be the matrix of first or second differences, or any other roughness operator. The $\boldsymbol{\beta}^{ols}$ is the ordinary least squares solution.

Name	Penalty term	Function definition
Ridge I	$\Psi = \lambda \sum_{i=1}^{p} g\left(\theta_i\right)$	$g(\theta) = \theta^2 ; \boldsymbol{\theta} = \boldsymbol{\beta}$
Ridge L		$g(\theta) = \theta^2 ; \boldsymbol{\theta} = \mathbf{L}\boldsymbol{\beta}$
LASSO		$g(\theta) = \lvert\theta\rvert ; \boldsymbol{\theta} = \boldsymbol{\beta}$
Fusion LASSO (FnLASSO)		$g(\theta) = \lvert\theta\rvert ; \boldsymbol{\theta} = \mathbf{L}\boldsymbol{\beta}$
Fused LASSO (FdLASSO)		$g^1(\theta) = \lvert\theta\rvert ; \boldsymbol{\theta}^1 = \boldsymbol{\beta}$ $g^2(\theta) = \lvert\theta\rvert \; \boldsymbol{\theta}^2 = \mathbf{\Omega}\boldsymbol{\beta}$
Smooth LASSO (SLASSO)	$\Psi = \lambda_1 \sum_{i=1}^{p} g^{(1)}\left\lvert\theta_i^{(1)}\right\rvert + \lambda_2 \sum_{i=1}^{p} g^{(2)}\left\lvert\theta_i^{(2)}\right\rvert$	$g^1(\theta) = \lvert\theta\rvert ; \boldsymbol{\theta}^1 = \boldsymbol{\beta}$ $g^2(\theta) = \theta^2 \; \boldsymbol{\theta}^2 = \mathbf{\Omega}\boldsymbol{\beta}$
Elastic Net L (ENET L)		$g^1(\theta) = \lvert\theta\rvert ; \boldsymbol{\theta}^1 = \mathbf{L}\boldsymbol{\beta}$ $g^2(\theta) = \theta^2 \; \boldsymbol{\theta}^2 = \mathbf{L}\boldsymbol{\beta}$
Adaptive LASSO (ALASSO)	$\Psi = \lambda \sum_{i=1}^{p} \gamma_i g\left(\theta_i\right)$	$g(\theta) = \lvert\theta\rvert ; \boldsymbol{\theta} = \boldsymbol{\beta}$ with $\gamma_i \geq 0$ for $i = 1,\ldots,p$
Nonnegative Garrote (NNG)		$g(\theta) = \lvert\theta\rvert ; \boldsymbol{\theta} = \boldsymbol{\beta}$ with $\gamma_i = 1 / \lvert\beta_i^{ols}\rvert$ for $i = 1,\ldots,p$ $\boldsymbol{\beta} \geq 0$ if $\boldsymbol{\beta}^{ols} > 0; \boldsymbol{\beta} \leq 0$ if $\boldsymbol{\beta}^{ols} < 0$

From the computational point of view, the local quadratic approximation (LQA) (Fan and Li, 2001) and the minorization-maximization (MM) (Hunter and Li, 2005) algorithms provide a numerical engine to implement several of the PLS methods. These algorithms are based on the Newton-Raphson (NR) technique, which uses a second-order Taylor approximation for the objective function to be minimized. In this sense, they inherit the virtues of the NR guaranteeing convergence at least to a local minimum, but they cannot deal with general MPLS models. To overcome this difficulty, Valdes-Sosa and colleagues introduced a generalized MM method for the estimation of massive autoregressive models of neuroimaging data (Valdés-Sosa et al., 2006), and for solving the M/EEG inverse problem (Vega-Hernández et al., 2008). Another drawback is that neither LQA nor MM produce true sparse solutions and depend on an additional parameter to assure numerical stability (Li et. al., 2006). Alternatively, the least angle regression (LARS) (Efron et al., 2004) and the Shooting algorithm (Fu, 1998; Friedman et al., 2007), offer efficient implementations for specific PLS methods (e.g., LASSO, ENET) with the advantage that they make variable selection and estimation simultaneously. Some other specific algorithms have been proposed to handle particular instances with specific difficulties, such as the lower bound based smoothed Quasi-Newton algorithm to compute "group bridge" solutions efficiently (Cao et al., 2016) but with an application scope less extensive than that of LQA and MM. Mørup et al. (2008) introduced a version of the LARS algorithm to implement the LASSO with nonnegative constraints and Chen et al. (2017) presented an improved-LARS algorithm, which combines LARS with Principal Component Analysis and Independent Weight evaluation. Chang et al. (2018) proposed a variant of the Adaptive LASSO (named Tukey-LASSO), with an accelerated proximal gradient algorithm which showed to provide better estimates as compared with the Adaptive LASSO and other robust LASSO implementations.

Section B. Quality Measures of the reconstruction.

Localization Error: Euclidean distance (in mm) between the maximum values of the estimated and real current densities: $Err = \mathbf{r}_{true} - \mathbf{r}$, where $\mathbf{r}_{true}$ and $\mathbf{r}$ denote the vector positions of the maximum of real and estimated PCD, respectively.

Normalized Localization Error: $$ErrN = 1 - \left(\frac{Err}{2R_b}\right)$$, where R_b is the radius of the brain (95 mm for the fitted sphere used in this work).

Blurring: Measure of the width or spatial dispersion of the solution. Here we compute the ratio between the full-width-at-half-maximum (FWHM) of the estimated and real current densities: $$Blurr = \frac{FWHM}{FWHM_{true}}$$. The FWHM was defined in Fuchs et al. (1999) as the volume of voxels with strength above 50% of the maximum PCD divided by the total volume of the brain (number of voxels times the volume of a voxel).

Normalized Blurring: $$BlurrN = \exp\left(2 - \left(Blurr + \left(\frac{1}{Blurr}\right)\right)\right)$$.

Visibility: The ratio between the maximum values of estimated and simulated PCD: $$Vis = \frac{\max\ \mathrm{PCD}}{\max\ \mathrm{PCD}_{true}}$$.

Normalized Visibility: $$VisN = \exp\left(2 - \left(Vis + \left(\frac{1}{Vis}\right)\right)\right)$$.